energies

MDPI

Article

Research on Object Detection of Overhead Transmission Lines Based on Optimized YOLOv5s †

Juping Gu [1,2,*], Junjie Hu [2], Ling Jiang [2], Zixu Wang [2], Xinsong Zhang [2], Yiming Xu [2], Jianhong Zhu [2] and Lurui Fang [3]

[1] School of Electrical and Information Engineering, Suzhou University of Science and Technology, Suzhou 215101, China
[2] School of Electrical Engineering, Nantong University, Nantong 226019, China
[3] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215028, China
* Correspondence: gu.jp@ntu.edu.cn
† This paper is an extended version of our paper published in 2022 the 12th International Conference on Power and Energy Systems (ICPES), Guangzhou, China, 23–25 December 2022; pp. 388–392.

Abstract: Object detection of overhead transmission lines is a solution for promoting inspection efficiency for power companies. However, aerial images contain many complex backgrounds and small objects, and traditional algorithms are incompetent in the identification of details of power transmission lines accurately. To address this problem, this paper develops an object detection method based on optimized You Only Look Once v5-small (YOLOv5s). This method is designed to be engineering-friendly, with the objective of maximal detection accuracy and computation simplicity. Firstly, to improve the detecting accuracy of small objects, a larger scale detection layer and jump connections are added to the network. Secondly, a self-attention mechanism is adopted to merge the feature relationships between spatial and channel dimensions, which could suppress the interference of complex backgrounds and boost the salience of objects. In addition, a small object enhanced Complete Intersection over Union (CIoU) is put forward as the loss function of the bounding box regression. This loss function could increase the derived loss for small objects automatically, thereby improving the detection of small objects. Furthermore, based on the scaling factors of batch-normalization layers, a pruning method is adopted to reduce the parameters and achieve a lightweight method. Finally, case studies are fulfilled by comparing the proposed method with classic YOLOv5s, which demonstrate that the detection accuracy is increased by 4%, the model size is reduced by 58%, and the detection speed is raised by 3.3%.

Keywords: overhead transmission line; object detection; larger scale detection layer; self-attention; bounding box regression; lightweight

Citation: Gu, J.; Hu, J.; Jiang, L.; Wang, Z.; Zhang, X.; Xu, Y.; Zhu, J.; Fang, L. Research on Object Detection of Overhead Transmission Lines Based on Optimized YOLOv5s. *Energies* **2023**, *16*, 2706. https://doi.org/10.3390/en16062706

Academic Editor: Georgios Christoforidis

Received: 14 February 2023
Revised: 5 March 2023
Accepted: 10 March 2023
Published: 14 March 2023

1. Introduction

Insulators and fittings are vital components of overhead transmission lines; their condition determines the safe operation of power systems. In their application, their working environment is harsh. No physical protection against the natural changes of the environment, such as wind and rain, determines a high probability of accelerating degradation and faults for insulators and fittings [1]. To understand their health state, routine inspection is normally adopted by power companies. Traditional manual inspection [2] requires field engineers to climb the tower to inspect the transmission line. With the increasing size of transmission lines, manual inspection methods are no longer a cost-effective solution. As an alternative, power companies adopt unmanned aerial vehicles (UAVs) for inspections [3]. UAVs would capture the images of transmission lines through equipped high-definition cameras. These would then send data, including the status of the transmission line equipment, to the ground station for processing.

Similar to all emerging technology, UAV power inspection brings new challenges because it requires engineers to evaluate and analyze the captured images [4]. With the increasing use of UAVs power inspection, the object detection demands for power companies shift from detecting large objects to detecting objects of different sizes. This incurs a problem of missing inspections if manual visual assessment still dominates. The efficiency will not be satisfactory as well. To address these problems, many investigations have developed object detection algorithms, which push UAV power inspection technology from offline detection to online detection [5]. Those methods can improve the efficiency of inspections and turns mass-scale application of UAV power inspection into a possibility.

From the technology development perspective, the existing object detection algorithms for transmission lines can be classified into two categories:

(1) Two-stage detection algorithms, such as Fast Region-Based Convolutional Neural Network (R-CNN) [6] and Faster R-CNN [7]. These algorithms would generate candidate regions before predicting the locations and classes of the objects in those regions. Among them, the literature [8] adopted the ImageNet dataset to pre-train the ResNET-101 network and realized the detection of insulators and bird nests. However, it was difficult to achieve real-time detection because of the high costs of the large capacity of memory. Study [9] utilized the improved Faster R-CNN for the detection of insulators and other components. This method improved the backbone with lightweight convolutional neural networks and added refinement modules at the output, which increased the detection speed without compromising detection accuracy. However, it did not focus on the interference of complex backgrounds in detection.

(2) One-stage detection algorithms, such as Single Shot Detector (SDD) [10] and YOLO [11], directly generate the class probabilities and position coordinates of the objects. Among them, [12] adopted improved SDD for insulators and spacers detection. This method used a mobile network to replace the original backbone. It also developed a deep feature pyramid to predict the output of different feature maps. Compared with Faster R-CNN, this method achieved better detection accuracy and faster speed. However, the feature collection capability for small objects was not satisfactory. Study [13] detected insulators using the improved YOLOv3. This method involved an additional spatial pyramid pooling module to merge local and global features. This method improved the detection effectiveness for objects of vastly different sizes. However, it compromised the detection accuracy for insulators within complex backgrounds. Study [14] investigated the detection of major electrical equipment using optimized YOLOv4, which developed a training strategy using CIoU and Generalized Intersection over Union (GIoU), respectively, for large and small objects. However, the classification of large and small objects was relatively not satisfactory in real applications. The authors of [15] simplified YOLOv4 by replacing the backbone network with the MobileNet, which reduced the model's complexity considerably and promoted its implementation generality on embedded systems. However, the simplification process sacrificed the detection accuracy. Study [16] used YOLOv5s to detect insulators, and added a transformer to the backbone to improve detection performance. The results showed that the detection speed of YOLOv5s is much faster than algorithms such as Faster R-CNN.

On the whole, of the two-stage algorithms, Faster R-CNN and others are not suitable for transmission line object detection due to large memory consumption and slow speed. Among the one-stage algorithms, YOLOv5 adopts the advantages of YOLOv3 and YOLOv4, and has better detection performance in theory. The case study [17] compared the detection of insulators by YOLOv3~v5. The case study justified that YOLOv5s has faster detection speed and smaller model size while having satisfactory detection accuracy. Although some scholars later proposed YOLOv6 and YOLOv7 on the basis of YOLOv5s, YOLOv5s is still one of the most widely used algorithms due to its simplicity and stability [18]. Drawing on the advantages of existing algorithms in the literature, YOLOv5s is chosen as the method for object detection of overhead transmission lines.

From the angle of the industrial application, the previous literature focuses on detecting only one piece or a small number of objects of overhead transmission lines. The gap between research and real application happened due to a lack of research on the scalability of object detection for overhead transmission lines. In detecting multi-class objects on overhead transmission lines, compared with objects of relatively apparent features, the available algorithms have a poor detection effect on objects with small sizes or complex backgrounds. This creates a data bottleneck and limits the improvement of detection accuracy. In addition, in previous research, a greater detection accuracy indicates higher model complexity and a lower possibility to be embedded on microdevices, and vice versa. There is no guarantee for a "win-win solution", i.e., improving detection accuracy and simplicity of detection algorithms at the same time.

To address the above problems, this paper develops an object detection method for overhead transmission lines based on optimized YOLOv5s. Firstly, the network structure is optimized by adding larger scale detection layers to retain more detailed features. Meanwhile, jump connections are introduced to achieve a balanced combination of multi-path features. Then, a self-attention mechanism is developed to combine the relationships between features from both spatial and channel dimensions into the feature map. Further, to automatically adjust loss for objects of different sizes, a small object enhanced CIoU loss is introduced as the loss function of the bounding box. Finally, L1 regularization is utilized to scale factors of the batch-normalization layers for sparse training. Channels are pruned according to the derived scale factors. Case studies justify that the proposed method can apply to object detection for transmission lines because of its high accuracy, small model size, and high speed.

2. YOLOv5s Principle

YOLOv5 [19] is an open object detection algorithm developed by the company Ultralytics. According to the complexity of the network, it includes several versions: Yolov5s, Yolov5m, Yolov5l, and Yolov5x. The structure of YOLOv5s is shown in Figure 1.

YOLOv5s mainly consists of four stages, which are introduced as follows.

(1) Input stage. YOLOv5s preprocesses the original image, including adaptive scaling, data enhancement, and the generation of initial anchors.

(2) Backbone stage. The main function is feature extraction, which consists of a focus module, convolution batch-normalization SiLU (CBS) module, C3 module, and spatial pyramid pooling (SPP) module [20]. The focus module is a special down-sampling operation that uses slicing operations to split a high-resolution feature map into multiple low-resolution feature maps. The CBS module is the most basic module in the network, consisting of a convolution layer, a batch-normalization layer, and a SiLU activation function. The C3 module is an efficient feature extraction module that can enhance the learning ability of the network. It consists of three 1×1 convolution layers and a bottleneck layer, hence the name C3 module. The bottleneck layer uses a 1×1 convolution layer to reduce dimension and a 3×3 convolution layer to extract features, which increases the depth of the network and reduces the amount of calculation. The SPP module uses different maximum pooling layers to convert feature maps of arbitrary size into a fixed size.

(3) Neck stage. This consists of a feature pyramid network (FPN) and path aggregation network (PAN) [21]. The FPN transmits high-level semantic information from top to bottom, while the PAN transmits shallow features from bottom to top. With this structure, feature maps of different scales are generated for detection.

(4) Output stage. This is responsible for predicting the coordinates, categories, and confidence scores of the objects. Moreover, it deletes invalid prediction results by non-maximum suppression (NMS) [22], and the final results are marked on the image.

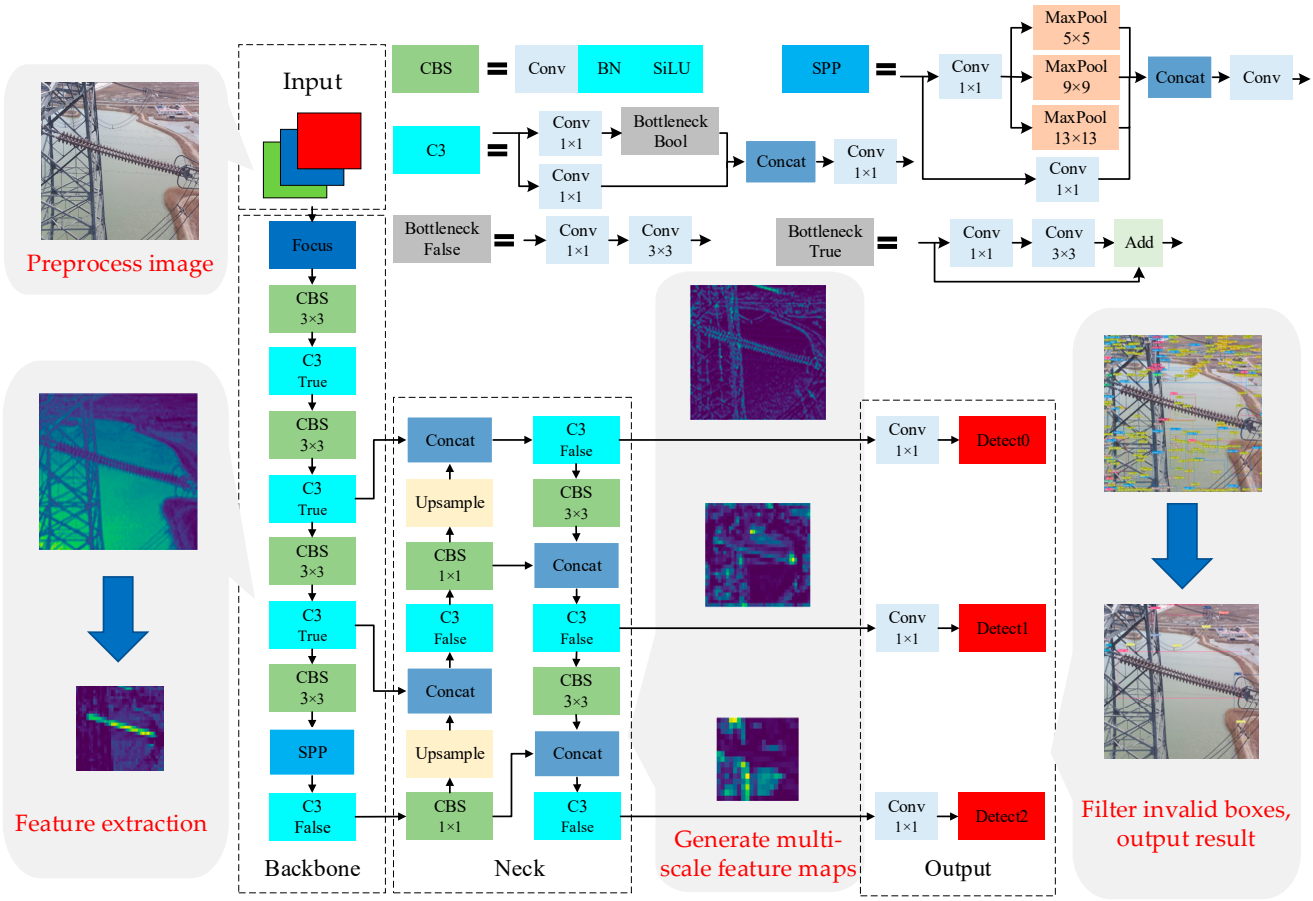

Figure 1. YOLOv5s structure.

3. Algorithm Improvement

This paper develops an object detection method for overhead transmission lines by optimizing the classic YOLOv5s. This method can address the detection problems of complex backgrounds and small objects in aerial images. Meanwhile, it adopts a lightweight method to improve computation efficiency, so as to promote the deployment simplicity.

3.1. Optimization of Network Structure

In capturing the images of overhead transmission lines, the UAV normally stages at a large distance away from the overhead lines. This causes the size of the part of captured components to be extremely small. This would incur missing detection of these objects, because of insufficient extractable features. Therefore, this paper optimizes the network structure to improve the UAV's ability to detect extremely small objects.

Firstly, taking extracted feature data from original images, YOLOv5s creates feature layers of 20 × 20, 40 × 40, and 80 × 80 pixels, where the 80 × 80 pixels feature layer has more detailed features and is used to detect small objects. However, with the development of camera technology, cameras could provide pictures with a much greater number of pixels than before. Therefore, adopting the 80 × 80 pixels feature map would cause the loss of small features during down-sampling. Increasing the scale of the feature layer is a solution to this problem. But if the feature layer is too large, network complexity will be greatly increased. To make a tradeoff between complexity and applicability, this paper adopts a feature layer of 160 × 160 pixels to detect small objects, as shown in the red background in Figure 2. The process is as follows: the 80 × 80 pixels feature map is upgraded to 160 × 160 pixels by C3, CBS, and upsample modules. Then, the upgraded feature map is connected with the 160 × 160 pixels feature map derived by the backbone stage to form

a new merged feature layer. Last, the merged feature layer is integrated by a C3 module for detection.

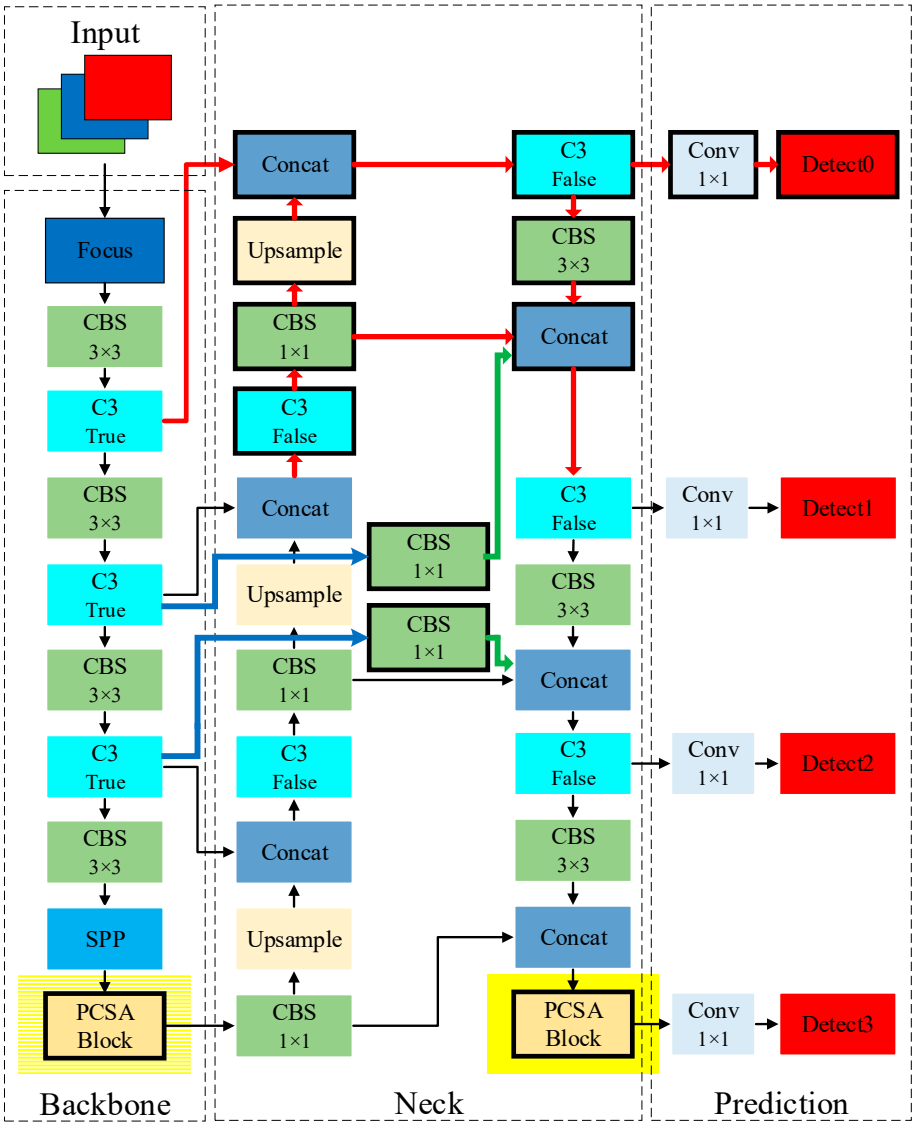

Figure 2. Optimized YOLOv5s structure.

In addition, in order to reduce the feature loss in the down-sampling process, jump connections are added between the backbone and neck stages to merge multi-path features. This improves the expression capability of the feature pyramid, as shown in the blue background in Figure 2. Meanwhile, since the number of channels in the backbone feature map is twice that of the neck, this paper utilizes 1×1 convolution modules to reduce the number of channels to balance the process of merging the multi-path features.

In order to verify the effectiveness of the optimization, we conducted a comparative experiment on the model before and after optimization. The results are shown in Figure 3. It can be seen that the original network misses the detection of some small objects, while the optimized network can accurately detect the adjusting plates and spacers, and can detect the occluded suspension clamp. It is verified that the above optimization methods can improve the extraction of detailed features of the network and can effectively improve the detection of small objects.

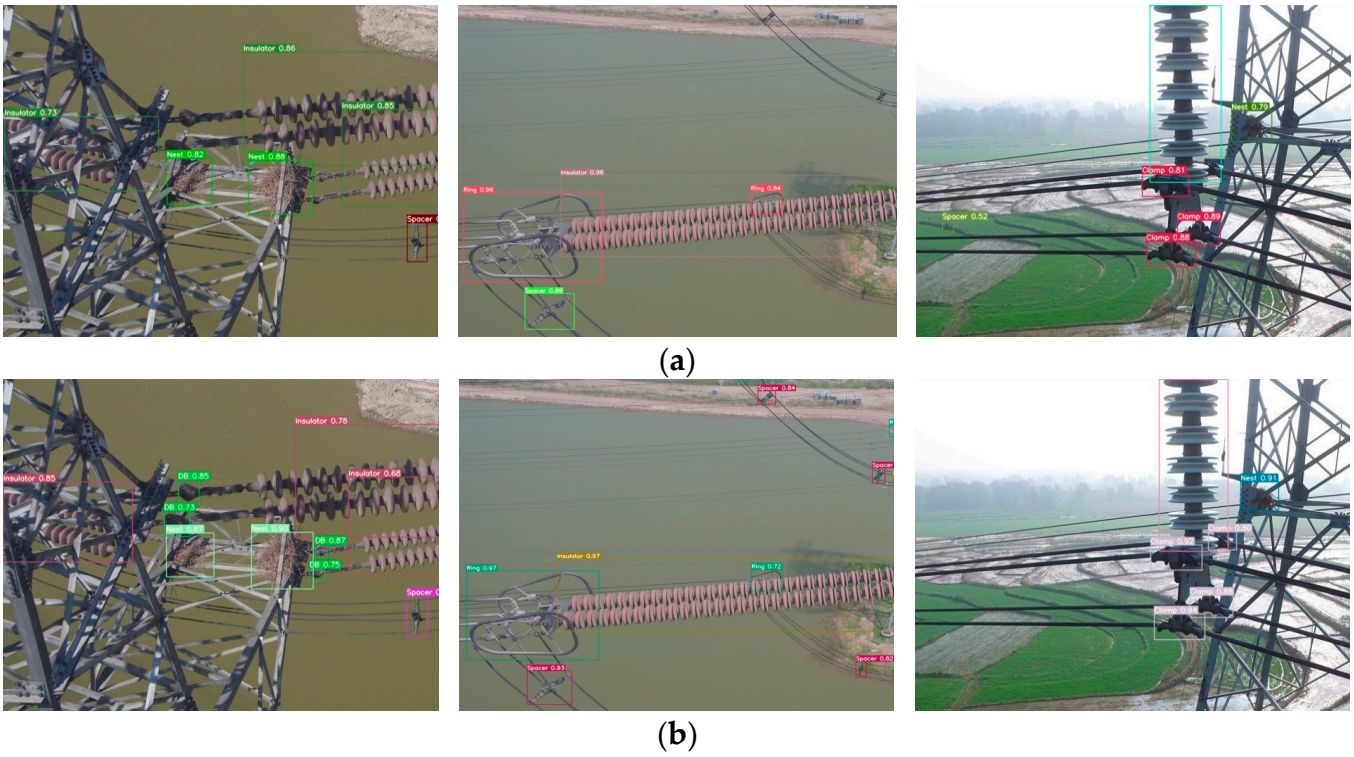

Figure 3. Comparative experiment. (**a**) Original network; (**b**) optimized network.

3.2. Addition of Self-Attention

To overcome the challenge of complex backgrounds of aerial images, this paper develops a self-attention mechanism called the position and channel self-attention (PCSA) block to replace the last C3 module of the backbone and neck stages, as shown in the yellow background in Figure 2. This self-attention mechanism could promote the significance of key features of objects, thus achieving improved accuracy.

The PCSA block still adopts the structure of C3 module and uses two different 1×1 convolution layers to reduce the input feature channels by half for processing, which can deepen the network while reducing the amount of calculation. The structure is shown in Figure 4.

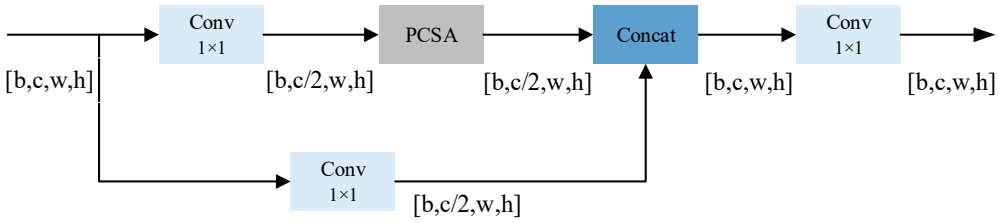

Figure 4. PCSA block structure.

The PCSA is proposed based on the non-local block [23]. It merges the position attention and the channel attention to make up for the insufficient attention of the non-local block. This structure is shown in Figure 5.

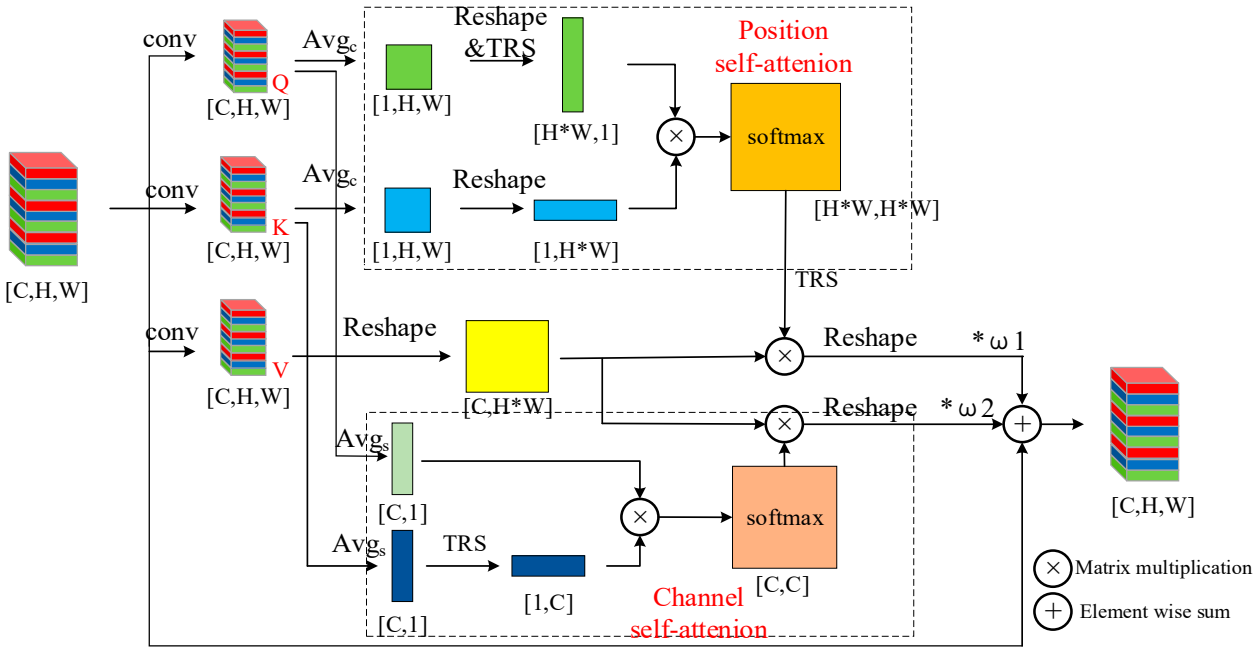

Figure 5. PCSA layer structure.

The position attention establishes the correlation between any two points in the spatial dimension as follows: (1) utilizing three different 1×1 convolutions to generate the corresponding feature maps Q, K, and V from original feature maps; (2) deriving the average value in channel dimension of Q and K; (3) multiplying Q and K to obtain a matrix of correlation; (4) adopting softmax to derive the position weights; and (5) multiplying V and the derived weight matrix to obtain the weighted feature map.1. This can be expressed as Equation (1).

The channel attention captures the relationship between different channels as follows: (1) Q and K are pooled on average in the spatial dimension; (2) they are matrix multiplied, followed by a softmax operation to obtain the channel weights; and (3) the channel weights and V are multiplied to obtain the weighted feature map.2. This can be expressed as Equation (2).

Finally, feature map.1 and feature map.2 are multiplied by the learnable parameters ω_1 and ω_2, respectively, and the original feature map X is summed to obtain the final feature map, as shown in Equation (3).

$$Out_1 = V\text{softmax}^{\mathrm{T}}(\text{avg}_c^{\mathrm{T}}(Q)\text{avg}_c(K)) \tag{1}$$

$$Out_2 = \text{softmax}(\text{avg}_s(Q)\text{avg}_s^{\mathrm{T}}(K))V \tag{2}$$

$$Out = \omega_1 Out_1 + \omega_2 Out_2 + X \tag{3}$$

where Q, K, V are the generated feature maps; avg_c is the average of the channel dimension; T is transposition; avg_s is the average pooling in the spatial dimension; ω_1 and ω_2 are learnable parameters; and X is the original feature map.

To verify the performance of the PCSA block module, a visualization experiment was carried out on it. Grad-CAM [24] is a visualization method that uses gradients to calculate the importance of features in convolutional layers, allowing the regions of interest to be clearly visible. The comparison results of the Grad-CAM heatmap after the addition of the PCSA module to the backbone network are shown in Figure 6.

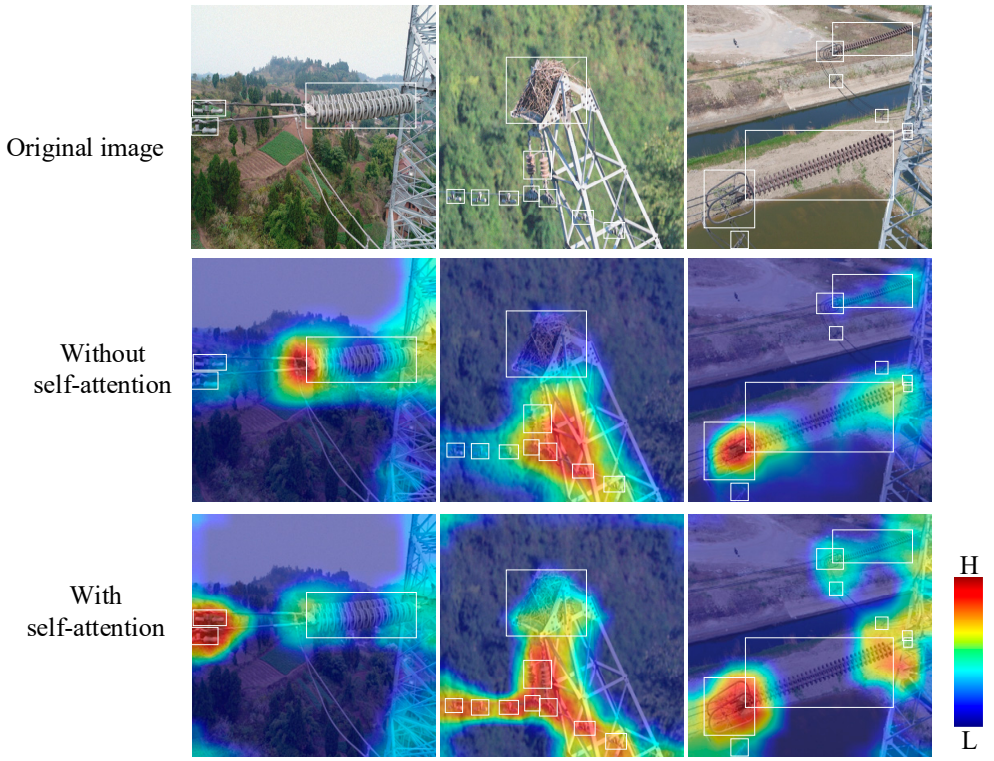

Figure 6. Heatmap experiment.

In Figure 6, the different colors of each area represent the different gradients between the current layer and the output layer. Red indicates that the area has a significant impact on the result. With the diminishing color of the area, the importance decreases. In the figure, we manually marked key objects with white boxes. By utilizing the PCSA module, we can see that the color of key objects goes deeper. This indicates that key areas become more distinguished compared to those without the PCSA model, for example, the dampers on the edge of the pictures in the first column, the bird nest in the middle of the pictures in the second column, and the insulator in the upper right of the pictures in the third column. The salience of these key objects has been enhanced, making it easier for the network to detect them. The improvement justified that using the PCSA model could effectively reduce the cases of missed and false detection, and improve the detection ability of objects under complex interference.

3.3. Optimization of Bounding Box Loss

Overhead transmission lines contain components of various sizes. In the process of bounding box regression, small objects are often more difficult to locate accurately due to their indistinct features. The detection effect of small objects restricts the overall detection effect. For the classic YOLOv5, the bounding box loss function uses CIoU loss (L_{CIoU}) [25]. However, it does not consider the influence of different areas. This paper proposes a small object enhanced CIoU loss (L_{SCIoU}) by adding the influence of the area factor. This function could automatically adjust the loss for different sizes of objects to improve the detection accuracy for small objects. Details are shown as follows:

$$\begin{cases} L_{SCIoU} = \frac{\lambda(1-IoU)+area^{gt}}{area^{gt}} L_{CIoU} \\ \qquad\quad = \frac{\lambda(1-IoU)+area^{gt}}{area^{gt}} (1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v) \\ \alpha = \frac{v}{(1-IoU)+v} \\ v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \end{cases} \quad (4)$$

where λ is the enhancement factor, adjusted by experiment; IoU is the intersection over union; $area^{gt}$ is the area of ground truth; b is the center point; ρ is Euclidean distance; c is the diagonal length of the smallest external rectangle of the two boxes; and h and w are the length and width of the box, respectively, as shown in Figure 7.

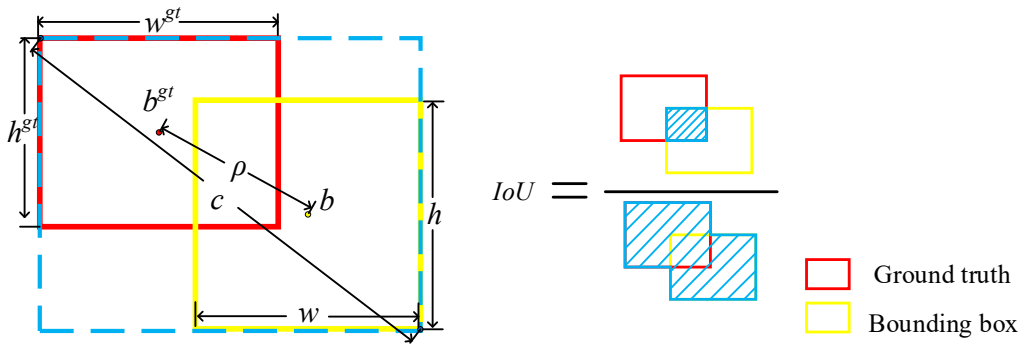

Figure 7. Parameter diagram.

Equation (4) indicates that $(1 - IoU)$ is a small value which is enhanced by λ. When $area^{gt}$ decreases, the values of $\frac{\lambda(1-IoU)+area^{gt}}{area^{gt}}$ and L_{SCIoU} increase. If $area^{gt}$ is diminishing, L_{SCIoU} increases more obviously. If $area^{gt}$ is larger, because $\lambda(1 - IoU)$ is much smaller than $area^{gt}$, the value of $\frac{\lambda(1-IoU)+area^{gt}}{area^{gt}}$ is roughly equal to 1. At this time, L_{SCIoU} is approximately equal to L_{CIoU}.

By increasing the bounding box loss of small objects, the weight of this part becomes larger, which motivates the network to pay more attention to small objects in the training phase and improve the regression accuracy correspondingly.

3.4. Reducing Computation Complexities

The classic YOLOv5s algorithm can achieve accurate detection on large servers. However, it does not perform well on mobile devices and other devices with insufficient GPU performance. To capture small objects, the network structure optimization in Section 3.1 increases the network complexity, which increases the deployment difficulty on less advanced devices. To promote the practical application of the proposed model, we employ a lightweight optimization based on the network slimming algorithm [26].

The YOLOv5s network is stacked by a large number of convolution modules. Each convolution module includes the convolution layer, batch-normalization (BN) layer, and SiLU activation function. Among them, the BN layer is used to speed up the training of the network and prevent gradient disappearance and network overfitting. The BN layer [27] normalizes the input data with a mean of 0 and a variance of 1, as given by:

$$\overline{x_i} = \frac{x_{in} - \mu}{\sqrt{\sigma^2 + \delta}} \tag{5}$$

$$y_{out} = \gamma \overline{x_i} + \beta \tag{6}$$

where x_{in} is the input data, μ is the mean of the input data, σ is the standard deviation of the input data, δ is a constant value, y_{out} is the output of the BN layer, γ is the scaling factor, and β is the translation factor.

According to Equation (6), the output of each channel in the BN layer is positively correlated to the scaling factor γ. If the value of γ is too small or close to 0, the channel output will remain very small and will have little effect on the detection results. At this time, the network complexity can be reduced by removing the convolution layer of its input and the channels of their output, as shown in Figure 8.

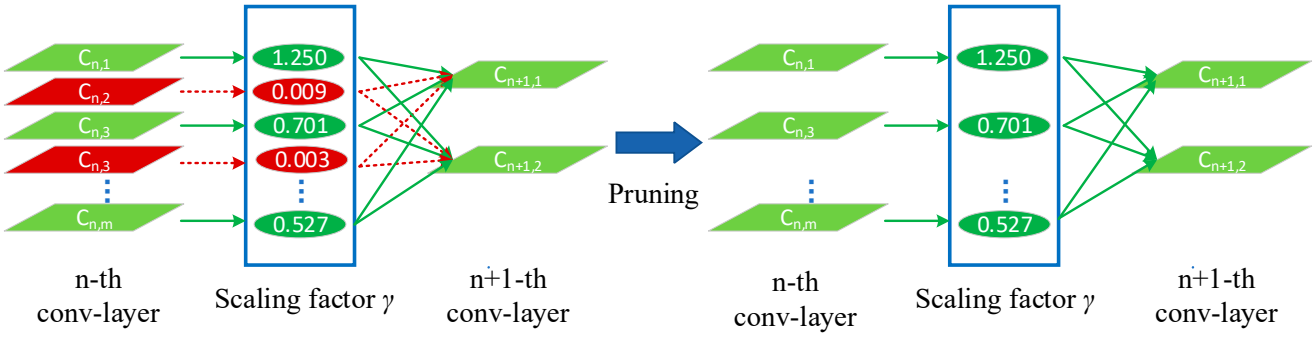

Figure 8. Pruning process.

However, under normal conditions, the distribution of scaling factor γ in the BN layer is close to the normal distribution in the training stage, as shown in Figure 9a. This makes it difficult to prune the network. L1 regularization is often used as a penalty term for the loss function in machine learning, which can produce a sparse weight matrix. Therefore, this paper adds the L1 regularization constraint to the γ value in the loss function of the BN layer to make sparse the network model, also known as sparse training, with the loss function as in Equation (7) [28].

$$L = \sum_{(a,b)} l(f(a,w),b) + \varphi \sum_{\gamma \in \eta} |\gamma| \tag{7}$$

$$\varphi = \theta(0.5\cos(\frac{epoch}{epochs}\pi) + 0.5) \tag{8}$$

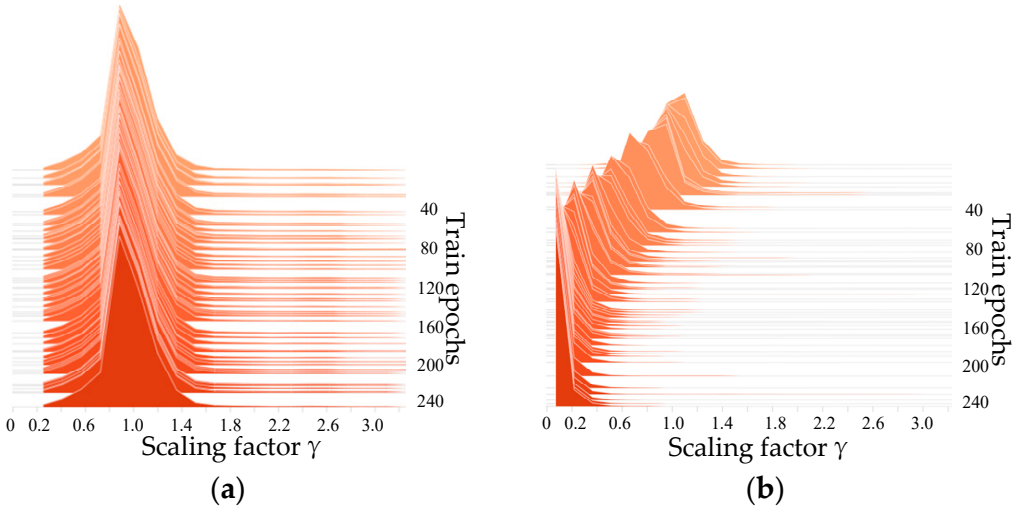

Figure 9. Histogram of scaling factor. (**a**) Normal training; (**b**) sparse training.

In Equation (7), $\sum l(f(a,w),b)$ represents the normal training loss function, $\varphi\sum|\gamma|$ represents the added L1 regularization constraint, a is the input, b is the target, w is the trainable weight, η is all the BN layer parameters, γ is the scaling factor, and φ is the sparsity parameter. If φ is too large, the network will lose too much accuracy, and if it is too small, the network will be insufficiently sparse. Equation (8) is proposed to adaptively reduce the sparsity parameters. At the beginning of training, a large sparsity parameter is given to make the network rapidly sparse. As the epoch increases, the sparsity parameter is reduced to compensate the accuracy. In Equation (8), θ is an adjustable parameter, determined by experiment; *epoch* is the current epoch; *epochs* is the total epoch.

After sparse training of the network, the values of some scaling factors γ in the BN layer converge to 0, as shown in Figure 9b. The values of this part have little effect on the

detection results. The network parameters can be reduced by removing the convolution layer of its input and the channels of their output. A global threshold can be set to adjust the pruning ratio according to the value ordering of all scaling factors γ. When the pruning ratio is high, it may result in accuracy loss, which can be partially recovered by fine-tuning the model. In this way, a more lightweight model can be obtained while maintaining accuracy.

4. Experiment and Analysis

4.1. Dataset Production

The dataset used in this paper consists of the public dataset of Chinese Power Line Insulators [29] and other real overhead transmission line aerial images, with a total of 1688 images. Most of the pictures were taken on alternating circuit transmission lines. The dataset involves various scenarios such as urban, rural, field, lake, and plain, including insulator, damper, shielding ring, spacer, counterweight, DB adjusting plate, grading ring, suspension clamp, sign, and bird nest, as shown in Figure 10.

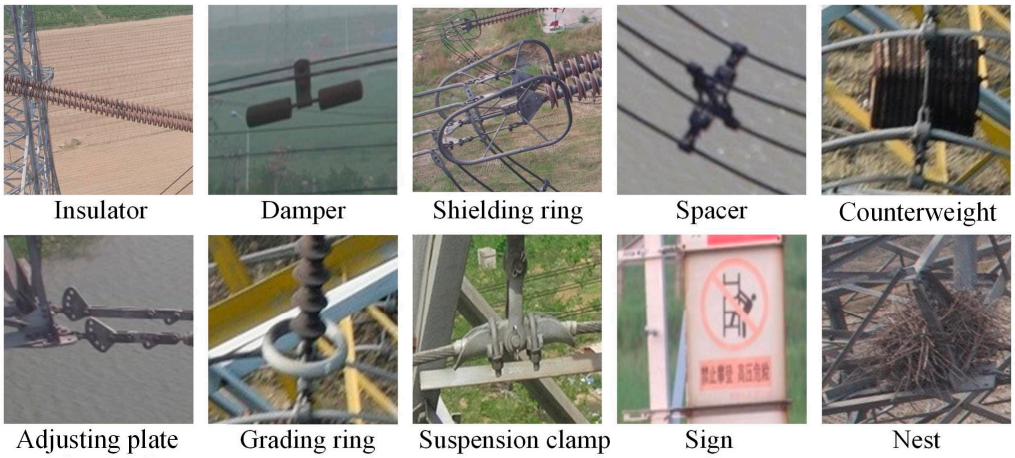

Figure 10. Objects to be detected.

Considering the insufficient number of images in the dataset, data enhancement methods were used to expand the images and improve the network generalizability. The data enhancement methods are shown in Figure 11. Each image needs to be randomly enhanced using two different methods. After data enhancement, 3376 images were generated. Training sets and test sets were divided in an 8:2 ratio.

4.2. Experimental Settings

The hardware and software parameters of the equipment used in the experiments are shown in Table 1.

The values of some important adjustable parameters involved in the experiments are as follows: the input image size is 640 $\times$ 640, the batch size is 16, the epoch is 250, the optimizer is stochastic gradient descent, the initial learning rate is 0.01, the final learning rate is 0.2, the momentum is 0.937, the enhancement factor λ is 20, and the sparsity training θ is 0.001.

4.3. Experimental Settings

In this paper, precision (P), recall (R), mean average precision (mAP) [30], and frame per second (FPS) are the main indicators for model performance, and the formulas are as follows:

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{9}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{10}$$

$$\mathrm{mAP} = \frac{\sum_{\mathrm{i}=1}^{N} AP_{\mathrm{i}}}{N} \tag{11}$$

where TP is the number of positive samples predicted to be positive; FP is the number of negative samples predicted to be positive; FN is the number of positive samples predicted to be negative; AP is the area enclosed by the P–R curve and the coordinate axis, which refers to the single-category accuracy rate; and N is the number of categories.

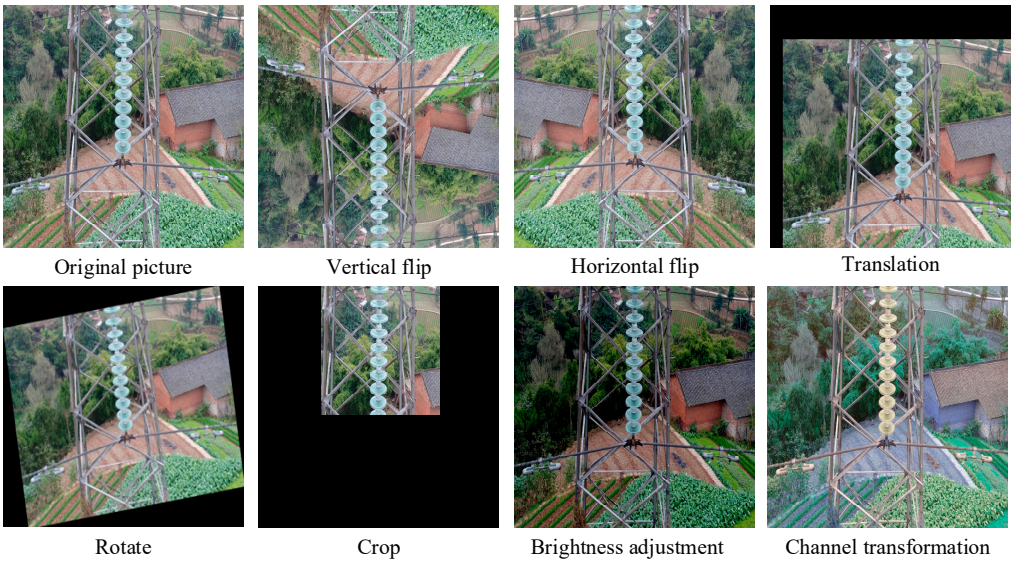

Figure 11. Data enhancement methods.

Table 1. Parameters of hardware and Software equipment.

Configuration	Parameters
CPU	Intel(R) Core (TM) i5-10300H CPU @ 2.50GHz
GPU	Nvidia GeForce RTX 1660ti SUPER 6GB
RAM	16GB
GPU acceleration library	Cuda11.0, Cudnn10.0

4.4. Experimental Results and Analysis

The comparison of the loss curves between YOLOv5s and the proposed method in the training stage is shown in Figure 12.

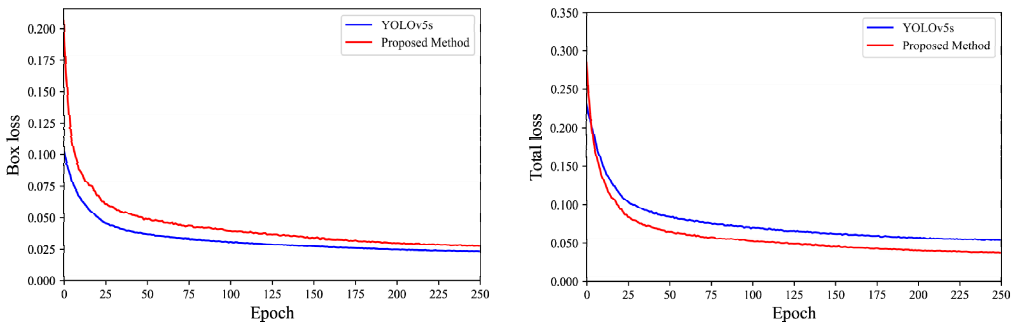

Figure 12. Loss curves. (**a**) Bounding box loss; (**b**) total loss.

In Figure 12, the initial bounding box loss of the proposed model is larger than that of YOLOv5s. This is because the L_{SCIoU} proposed in this paper increases the bounding box loss

of small objects. However, with the increase in the number of epochs, the final convergence value is almost the same as that of YOLOv5s. In addition, compared to YOLOv5s, the total loss of the proposed model decreases faster, and the convergence value is lower. This indicates that the training effectiveness of the proposed model is better.

After the training phase, the best weight of each model (unpruned) was selected for testing. Table 2 presents the comparison of detection accuracy between YOLOv5s and the proposed method. Compared to YOLOv5s, the proposed method improves the average detection accuracy for each type of object. Among them, the mAP of the adjustment plate increased by 6.9%, and that of suspension clamp increased by 10.6%; both of these sizes are relatively small, indicating that the proposed method can improve the detection performance of small objects. The total mAP of the proposed model exceeds that of YOLOv5s by 4.2%.

Table 2. Comparison of detection results in the test set.

	YOLOv5s			Proposed Method		
Category	**P (%)**	**R (%)**	**mAP@0.5 (%)**	**P (%)**	**R (%)**	**mAP@0.5 (%)**
Insulator	95.7	93.1	94.6	95.9	93.6	96.1
Damper	93.3	86.6	89.5	93.4	88.9	92.2
Adjusting plate	92.5	85.1	88.9	95.1	91.0	95.1
Spacer	95.0	94.0	95.2	95.8	95.1	96.9
Shielding ring	97.3	95.0	96.4	97.4	96.0	96.9
Grading ring	92.1	83.5	86.6	95.7	89.4	91.9
Counterweight	94.8	84.4	87.5	96.8	87.3	90.9
Sign	92.3	78.6	84.3	92.6	89.4	91.8
Suspension clamp	86.0	80.7	83.9	94.1	86.9	92.8
Bird nest	93.3	91.6	92.0	93.4	92.0	92.3
All	93.2	87.3	89.9	95.0	91.0	93.7

To analyze the lightweight method, pruning experiments were carried out on the model. The above best weight was pruned and fine-tuned after the sparse training stage. The sparse model was pruned in different ratios, and the experimental results are shown in Table 3. It presents, with a 20% pruning ratio or less, the mAP of the network remains. This indicates that the deleted channel is correctly selected. Pruning these channels would not influence the model's performance. The mAP of the model slightly reduces when the pruning ratio is 30% and 40%. This decrease is acceptable, compared to the decrease in the number of parameters and the floating point of operations (FLOPs). However, the mAP of the model drops sharply when the pruning ratio is above 40%. Under this circumstance, a part of the channels with high weights is pruned. They perform vital functions in feature extraction. Therefore, to minimize the model complexity as much as possible while maintaining the model accuracy, the model with a pruning ratio of 40% is selected as the benchmark for further comparison in this paper.

Table 3. Comparison of models with different pruning ratios.

Pruning Ratio	mAP@0.5 (%)	Parameters (M)	FLOPs (G)	FPS	Model Size (MB)
0%	93.7	6.4	19.3	49	12.5
10%	93.7	5.5	18.0	52	11.0
20%	93.7	4.7	17.0	55	9.3
30%	93.6	3.6	15.8	60	7.5
40%	**93.5**	**2.7**	**14.6**	**63**	**5.7**
50%	88.1	2.1	13.3	68	4.6
60%	74.6	1.6	11.9	72	3.4

After the pruning experiments, the proposed model was tested on real images. Several typical images are selected for display, as shown in Figure 13. It can be seen that when the picture is filled with objects of different sizes, the proposed model is basically able to

detect them. At the same time, the proposed model has a high confidence in detecting small objects such as a suspension clamp and adjusting plate. In addition, the proposed model has a strong ability to detect objects in complex backgrounds.

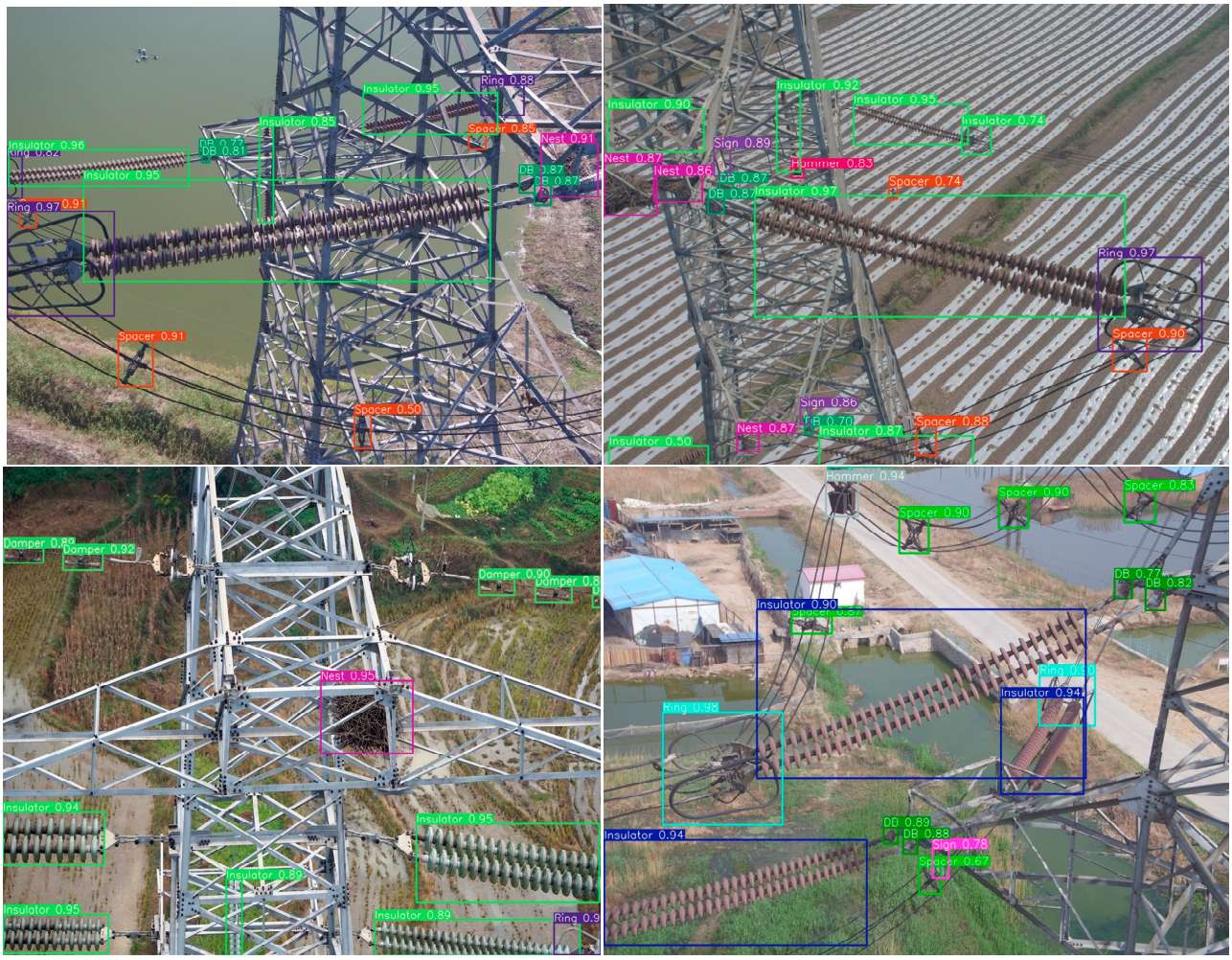

Figure 13. Detection results on pictures.

To further investigate the effects of the improved methods proposed in this paper on various aspects of the model, ablation experiments were conducted. The results are shown in Table 4, and the '√' represents the adoption of this method. The improvement of the network structure significantly improves the mAP but also increases the parameters and FLOPs, and reduces the speed. Using self-attention to replace the C3 module improves the mAP while reducing the parameters and FLOPs. Using L_{SCIoU} improves the mAP slightly without changing others. Pruning (40%) can greatly reduce parameters and FLOPs at a slight sacrifice of the mAP.

To justify the effectiveness of the proposed model, this paper compares the detection accuracy and computation simplicity of the proposed model with those of other state-of-the-art methods. The comparison studies are shown in Table 5. Compared with the traditional algorithm Faster R-CNN, the proposed method shows a significant improvement in mAP and speed. Compared with the latest algorithm YOLOv7, the proposed method has a faster speed under the same mAP. Overall, the proposed model has 2.7M parameters and a 5.7MB model size, which is significantly less than all other algorithms. This implies that the model size is small and easy to be applied. In addition, the detection speed is faster than other algorithms, indicating that it can better meet the requirement for real-time detection. Most importantly, the mAP of the proposed model is greater than other algorithms, indicating

that the detection effect is better than other algorithms. In detail, compared to the classic YOLOv5s, the mAP is improved by 4%, the model size is reduced by 58.4%, and the speed is improved by 3.3%.

Table 4. Ablation experiments.

Structure Improvement	Self-Attention	Box Loss	Pruning (40%)	Parameters (M)	FLOPs (G)	Model Size (MB)	FPS	mAP@0.5 (%)
				7.1	16.4	13.7	61	89.9
√				7.3	19.7	14.3	49	91.6
	√			6.2	16.0	12.0	60	91.1
		√		7.1	16.4	13.7	61	90.8
	√	√		6.2	16.0	12.0	60	91.5
√		√		7.3	19.7	14.3	49	92.0
√	√			6.4	19.3	12.5	49	92.9
√	√	√		6.4	19.3	12.5	49	93.7
√	√	√	√	2.7	14.6	5.7	63	93.5

Table 5. Comparison of mainstream algorithms.

Algorithm	Image Size (PX)	Parameters (M)	Model Size (MB)	FPS	mAP@0.5 (%)
Faster R-CNN	600 × 600	137.1	522	10	74.6
CenterNet	640 × 640	32.7	124	30	79.6
YOLOv3	640 × 640	61.9	235	19	88.0
YOLOv4	640 × 640	64.4	244	20	89.9
YOLOXs	640 × 640	8.9	34.3	56	89.1
YOLOv5s	640 × 640	7.1	13.7	61	89.9
YOLOv5m	640 × 640	21.8	42.5	19	90.3
YOLOv5l	640 × 640	46.7	89.4	14	90.5
YOLOv5x	640 × 640	87.3	167.0	5	91.2
YOLOv7	640 × 640	37.2	71.4	9	93.4
Proposed Method	640 × 640	2.7	5.7	63	93.5

Figure 14 shows comparison studies on images, with the following observations: (1) Faster R-CNN and CenterNet have poor overall detection performance. They are disturbed by the background of the iron tower to cause false detection and cannot accurately detect small-sized adjusting plates and dampers; (2) YOLOv4 has high confidence in the detection of various objects, but it fails to detect the adjustment plates; (3) YOLOv5s fails to detect bird nests in complex backgrounds, as well as small-sized adjusting plates and dampers. Compared with the classic YOLOv5s, the proposed model greatly improves its detection effect on both small objects and objects in complex backgrounds. Therefore, the proposed model is more suitable for transmission line object detection.

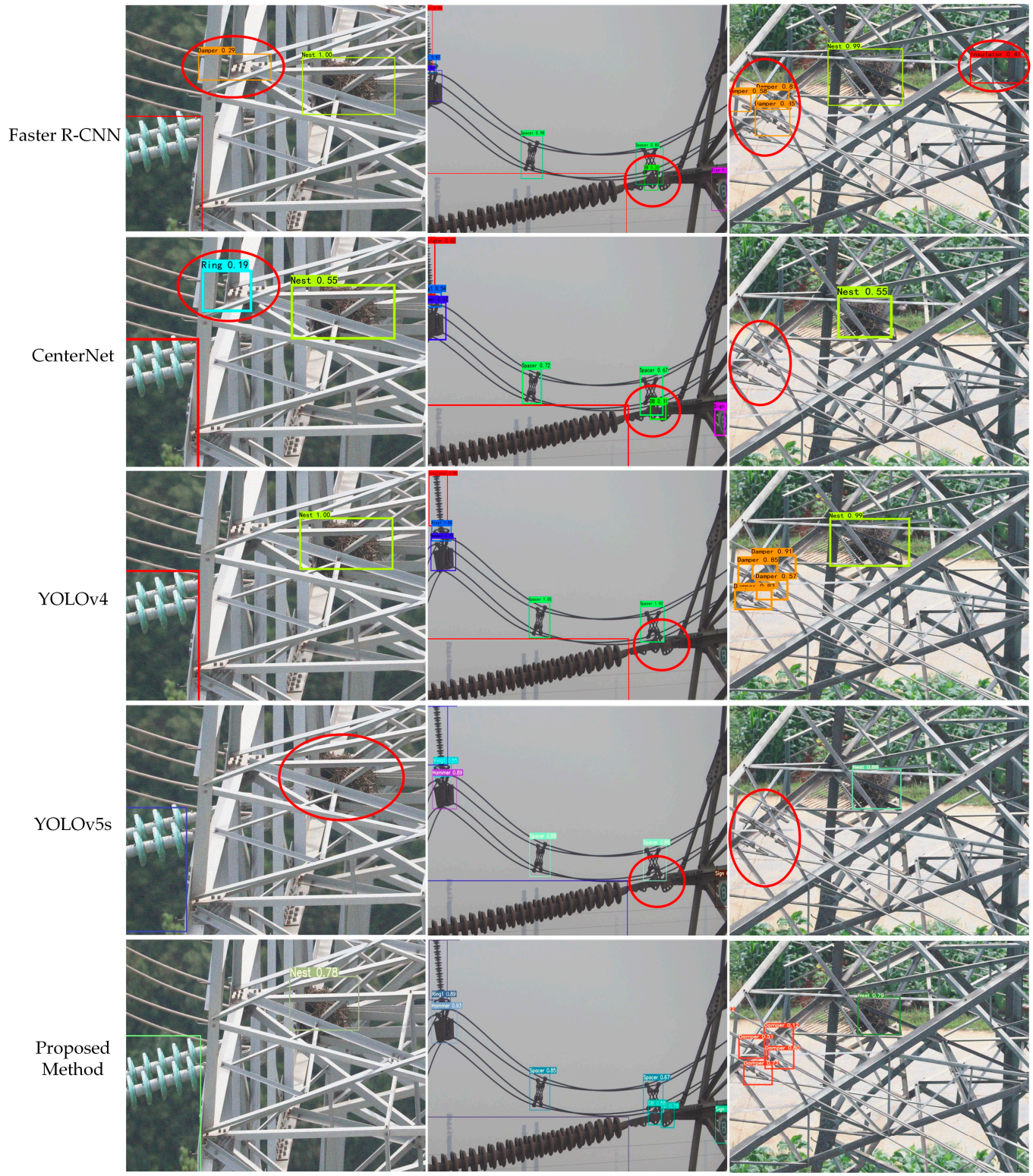

Figure 14. Comparison of detection results of mainstream algorithms.

5. Conclusions

This paper develops an object detection method for overhead transmission lines based on optimized YOLOv5s. The method improves the detection of small objects by improving the network structure and utilizing a L_{SCIoU} box loss function. Then, a self-attention mechanism is adopted to suppress the interference of complex backgrounds. Further,

channel pruning based on BN layers is performed to realize a lightweight model. Case studies justify that, compared to classic YOLOv5s, the mAP of the proposed model is improved by 4%, the model size is reduced by 58%, and the detection speed rises by 3.3%. These achievements are vital in real applications, therefore promoting industry implementation in UAV-based transmission line inspections.

Author Contributions: Conceptualization, J.G.; methodology, J.G. and J.H.; software, J.H.; validation, L.J.; formal analysis, Z.W.; investigation, X.Z.; resources, Y.X.; data curation, J.Z.; writing—original draft preparation, J.H.; writing—review and editing, J.G. and L.F.; supervision, J.Z.; project administration, J.G.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Program of National Natural Science Foundation of China, grant number U2066203; the Key Research and Development Plan of Jiangsu Province, grant number BE2021063; the Natural Science Foundation of the Jiangsu Higher Education Institutions of China, grant number 21KJA470006 and 22KJA470006.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to thank the editor and reviewers for their suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gu, J.; Hu, J.; Jiang, L. Object detection of overhead transmission lines based on improved YOLOv5s. In Proceedings of the 12th International Conference on Power and Energy Systems (ICPES), Guangzhou, China, 23–25 December 2022; pp. 388–392.
2. Yang, L.; Fan, J.; Liu, Y. A review on state-of-the-art power line inspection techniques. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9350–9365. [CrossRef]
3. Nguyen, V.; Jenssen, R.; Roverso, D. Intelligent monitoring and Inspection of power line components powered by UAVs and deep learning. *IEEE Power Energy Technol. Syst. J.* **2019**, *6*, 11–21. [CrossRef]
4. Manninen, H.; Ramlal, C.; Singh, A. Multi-stage deep learning networks for automated assessment of electricity transmission infrastructure using fly-by images. *Electr. Power Syst. Res.* **2022**, *209*, 107948. [CrossRef]
5. Kim, S.; Kim, D.; Jeong, S. Fault diagnosis of power transmission lines using a UAV-mounted smart inspection System. *IEEE Access* **2020**, *8*, 149999–150009. [CrossRef]
6. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 1440–1448.
7. Li, F.; Xin, J.; Chen, T. An automatic detection method of bird's nest on transmission line tower based on Faster-RCNN. *IEEE Access* **2020**, *8*, 8164214–8164221.
8. Lei, X.; Sui, Z. Intelligent fault detection of high voltage line based on the faster R-CNN. *Measurement* **2019**, *138*, 379–385. [CrossRef]
9. Wei, Y.; Li, M.; Xie, Y. Transmission line inspection image detection based on improved Faster-RCNN. *Electr. Power Eng. Technol.* **2022**, *41*, 171–178.
10. Liu, W.; Anguelov, D.; Erhan, D. SSD: Single shot multibox detector. In Proceedings of the 2016 European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016; pp. 21–37.
11. Redmon, J.; Divvala, S.; Girshick, R. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
12. Liu, X.; Li, Y.; Shuang, F. ISSD: Improved SSD for insulator and spacer online detection based on UAV system. *Sensors* **2020**, *20*, 6961. [CrossRef] [PubMed]
13. Tang, R.; Zhang, M.; Xu, H. Insulator recognition in transmission line inspection images based on deep learning. *Power Grid Clean Energy* **2020**, *37*, 41–46.
14. Li, F.; Niu, L.; Wang, S. Intelligent detection and parameter adjustment strategy of main electrical equipment based on optimized YOLOv4. *Trans. Chin. Electrotechn. Soc.* **2021**, *36*, 4837–4848.
15. Qiu, Z.; Zhu, X.; Liao, C. Detection of Transmission Line Insulator Defects Based on an Improved Lightweight YOLOv4 Model. *Appl. Sci.* **2022**, *12*, 1207. [CrossRef]
16. Yin, T.; Liang, J.; Liang, X. An Improved Model Based on Deep Learning for Detecting Insulator Defects. In Proceedings of the 2022 5th International Conference on Information Communication and Signal Processing (ICICSP), Shenzhen, China, 16–18 September 2022; pp. 229–233.
17. Liu, D.; Deng, C.; Zhang, H. Adaptive Reflection Detection and Control Strategy of Pointer Meters Based on YOLOv5s. *Sensors* **2023**, *23*, 2562. [CrossRef] [PubMed]

18. Li, Q.; Zhao, F.; Xu, Z. Insulator and damage detection and location based on YOLOv5. In Proceedings of the 2022 International Conference on Power Energy Systems and Applications (ICoPESA), Singapore, 25–27 February 2022; pp. 17–24.
19. Danso, S.; Shang, L.; Hu, D. Hidden Dangerous Object Recognition in Terahertz Images Using Deep Learning Methods. *Appl. Sci.* **2022**, *12*, 7354. [CrossRef]
20. Ding, J.; Cao, H.; Ding, X. High Accuracy Real Time Insulator String Defect Detection Method based on Improved YOLOv5. *Front. Energy Res.* **2022**, *10*, 889. [CrossRef]
21. Liu, J.; Liu, C.; Wu, Y. An Improved Method Based on Deep Learning for Insulator Fault Detection in Diverse Aerial Images. *Energies* **2021**, *14*, 4365. [CrossRef]
22. Zhao, Z.; Zhen, Z.; Zhang, L. Insulator Detection Method in Inspection Image Based on Improved Faster R-CNN. *Energies* **2019**, *12*, 1204. [CrossRef]
23. Wang, X.; Girshick, R.; Gupta, A. Non-local neural networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–21 June 2018; pp. 7794–7803.
24. Selvaraju, R.; Cogswell, M.; Das, A. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 618–626.
25. Zheng, Z.; Wang, P.; Liu, W. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the 2020 AAAI Conference on Artificial Intelligence (AAAI), New York, NY, USA, 7–12 February 2020; pp. 4322–4323.
26. Zhuang, L.; Li, J.; Shen, Z. Learning Efficient Convolutional Networks through Network Slimming. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 24–27 October 2017; pp. 2755–2763.
27. Kalayeh, M.; Shah, M. Training Faster by Separating Modes of Variation in Batch-Normalized Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 1483–1500. [CrossRef] [PubMed]
28. Chen, S.; Zhan, R.; Wang, W. Learning Slimming SAR Ship Object Detector Through Network Pruning and Knowledge Distillation. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **2021**, *14*, 1267–1282. [CrossRef]
29. Tao, X.; Zhang, D.; Wang, Z. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 1486–1498. [CrossRef]
30. Lu, X.; Zhang, Y.; Yuan, Y. Gated and Axis-Concentrated Localization Network for Remote Sensing Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 179–192. [CrossRef]